Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 171
DOI: 10.11159/icsta24.171

Development of LSTM Based Models for Air Pollutant-Related Public Health Effects

Huawei Han, Wesley S. Burr
Trent University
1600 West Bank Drive, Peterborough, ON Canada
rebeccahan@trentu.ca; wesleyburr@trentu.ca

Abstract **–** Air pollutants are considered to pose significant risk for public health and are often taken as one of the major concerns in related environmental epidemiology studies. Various statistical methods have been developed to assess the impact of short-term air pollutants exposure on human health, with Generalized Additive Models (GAMs) being the most widely-used models for their health risk response interpretability. However, challenges still exist for GAMs when dealing with multiple air pollutants as well as assessing health outcomes from accumulated exposure impacts with distributed lags. Considering the advancement of neural networks in recent years, this paper proposes a long short-term memory (LSTM) architecture-based model for air pollutant-related public health effect assessment. Datasets from the National Morbidity, Mortality and Air Pollution Study (NMMAPS) program are first prepared, and then an LSTM based health effect model with weighted evaluation of impacts from exposure to air pollutants with distributed lags is presented. Test results show that the proposed model has great potential in assessing the influence of air pollutants on public health effects, taking advantage of accumulative lagging impacts of multiple air pollutants exposure.

Keywords**:** air pollutants, human health, distributed lags, long short-term memory, environmental epidemiology

1. Introduction

Short- and long-term exposure to ambient air pollutants is considered to be one of the major concerns in environmental epidemiology studies for their adverse effects on human health. Health effects including premature mortality, morbidity, hospitalization, pulmonary diseases and cardiovascular diseases are considered to be connected with typical air pollutants such as particulate matter (PM), and oxides of sulphur (SO_x), ozone (O_3) and nitrogen (NO_x) [1-2]. In order to quantitatively evaluate their short-term adverse effects on human health, statistical methods like generalized linear models (GLMs) or GAMs are extensively used in this health assessment problem [3-9]. As people are usually exposed to multiple air pollutants simultaneously, research have been conducted to extend from investigating the influence of a single air pollutant (in earlier work) to exploring the influence of multiple such pollutants. The adverse effects of multiple air pollutants on human health are still a research hotspot today, e.g., [10-11]. However, challenges still exist when evaluating the combined influence of multiple air pollutant exposure on human health. It is not easy to design GLMs or GAMs in order to effectively capture the relation between a specific health effect and exposures to air pollutants of interest when considering distributed lags, especially when dealing with the uncertain confounding effects and collinearity – due to constraints on data collection and the public health design, individual exposures are not available, and the interactions between the pollutant predictors are unknown. Additionally, GAM-type health assessment models used in epidemiology study generally aim to investigate the form of the health outcome response to the air pollutants' exposure variation to assist in policy making. The assessment of their fitting performance or reliability of parameters are typically made using the same datasets the data are fit to.

LSTM models are improved versions of recurrent neural networks (RNNs) and were developed to deal with chronological dependence in machine learning problems with sequenced inputs [12]. Its recurrent structure has shown superior performance in natural language processing [13], speech recognition [14], and time series forecasting [15], among many fields. As some public health effects of interest are generally considered to be related to accumulated short-term exposure to air pollutants, this work proposes an LSTM network-based model for assessment of adverse impacts of ambient air pollutants on human health. Besides the advantage of effectively handling sequenced air pollutants exposures, the LSTM network-based model also provides an effective structure in capturing the joint effects of multiple air pollutants on health outcomes. In this work, these two advantages are the main inspirations for applying LSTM architecture to the air pollutants-

related public health evaluation problem in order to deal with the aforementioned issues of the traditional GAMs framework – including multiple nonlinearly-related air pollutant predictors in a model simultaneously. The rest of this paper is organized as follows. In Section 2, the air pollutants datasets from NMMAPS program are pre-processed and prepared for model training and testing. In Section 3, an LSTM network-based public health effect model is proposed with weighted evaluation of output features from LSTM network, which assesses the impacts of exposure to air pollutants with distributed lags on human health. In Section 4, the performance of the proposed model is tested and compared with different lags of exposure. Finally, a brief conclusion is made in Section 5.

2. Data Preparation

The air pollutants and health consequence data used in this work are from the National Morbidity, Mortality and Air Pollution Study program funded by the Health Effects Institute of the United States [16]. One of its aims was to study the association between air pollutants and daily mortality in large cities. Datasets of Chicago from Jan. 1, 1987 to Dec. 31, 2000 are used in the following analysis, which include measurement of six major regulated air pollutants (SO_2, NO_2, CO, O_3, PM_{10} and $PM_{2.5}$) on a daily basis and daily non-accidental mortality. Temperature is also considered as one of the factors that is related to the health outcome. In order to facilitate model training and analysis, the original data are pre-processed and prepared as follows. First, missing values of air pollutants (daily mean) are interpolated [17]. Second, a few highly extreme outliers in daily mortality and daily mean of each air pollutant are replaced with corresponding average value of days before and after the day with anomaly (e.g., spike mortality occurs on Jul. 15, 1995 due to an extreme heat wave, and is replaced with the average of mortalities in Jul. 14 and Jul. 16 of 1995). Then the original data are rescaled:

$$x_{\mathrm{s},i} = \frac{x_i - \mu_x}{\sigma_x} \tag{1}$$

where x_i is the i-th value in the original dataset for each category, μ_x and σ_x are their mean and standard deviation, and $x_{s,i}$ is the corresponding standardized value. After standardization, each category of original data is rescaled to have a mean value of 0 and a standard deviation of 1. The main purpose of standardizing the original data is to eliminate the influence of dimension for each category as the ambient air pollutants are generally measured and recorded with different metrics and have different scales. The other reason for using standardization to pre-process the data is to accelerate the solving process of searching for the optimal parameters during the training of model [18]. Finally, the datasets are reorganized with each element being a series of the original data. For example, assume the length of each element of a specific category of air pollutant data after reorganization is m, then the reorganized data sample is shown as Eqn. (2),

$$\boldsymbol{x}_{\mathrm{new},i} = [x_{\mathrm{s},i-m+1}, x_{\mathrm{s},i-m+2}, \ldots, x_{\mathrm{s},i}] \tag{2}$$

where $x_{\mathrm{new},i}$ is the i-th reorganized data sample with length m. The new data series becomes $[\boldsymbol{x}_{\mathrm{new},1}, \boldsymbol{x}_{\mathrm{new},2}, \ldots, \boldsymbol{x}_{\mathrm{new},n-m+1}]$ with n being the number of original data samples. The main purpose of this reformulation of the air pollutants data is to facilitate the model training process. The length m is the same as the time lag used in GAM-type health effect model, while the difference is that Eqn. (2) has distributed lags and the GAMs generally have a single lag. Thus this model can be thought of as combining some of the philosophy of the traditional model and the work on distributed lag models, e.g., DLNMs [22] or synthetic lag [23].

3. LSTM Based Public Health Consequence Assessment Model

From the world of neural networks and machine learning, LSTM is a form of improved RNN that is designed to deal with sequenced data [19-20]. By handling the gradient vanishing problem and gradient explosion problem using gate mechanisms, LSTM can deal with long-term dependency of data sequence and extract information from periods of time effectively. Its basic cell has several specially designed neural network layers interacting with each other to control the information flow and make decisions on what information should go through, be stored in the network or be filtered out. Through manipulation of information from a data sequence in each cell and repeating of these basic cells, a LSTM network is constructed and is capable of extracting features and information from a data sequence for analysis.

In air pollutants-related public health studies, the health outcome (e.g., mortality, morbidity) is generally assumed to be influenced by a short timescale of exposure to specific air pollutants [2-8, 21]. That is to say, the health outcomes are temporally related to historical exposures and are results of accumulated impacts. However, the impacts of exposure to air pollutants with distributed lags, specifically daily exposures before the consequence occurs, is generally not assessed and included in the GAM-type health consequence assessment models. Considering these issues and the aforementioned merits of LSTM framework, an LSTM based model with weighted evaluation of adverse impacts of historical exposure is proposed and applied to air pollutants-related public health consequence assessment. The structure of the proposed model is shown in Fig. 1 and detailed as follows.

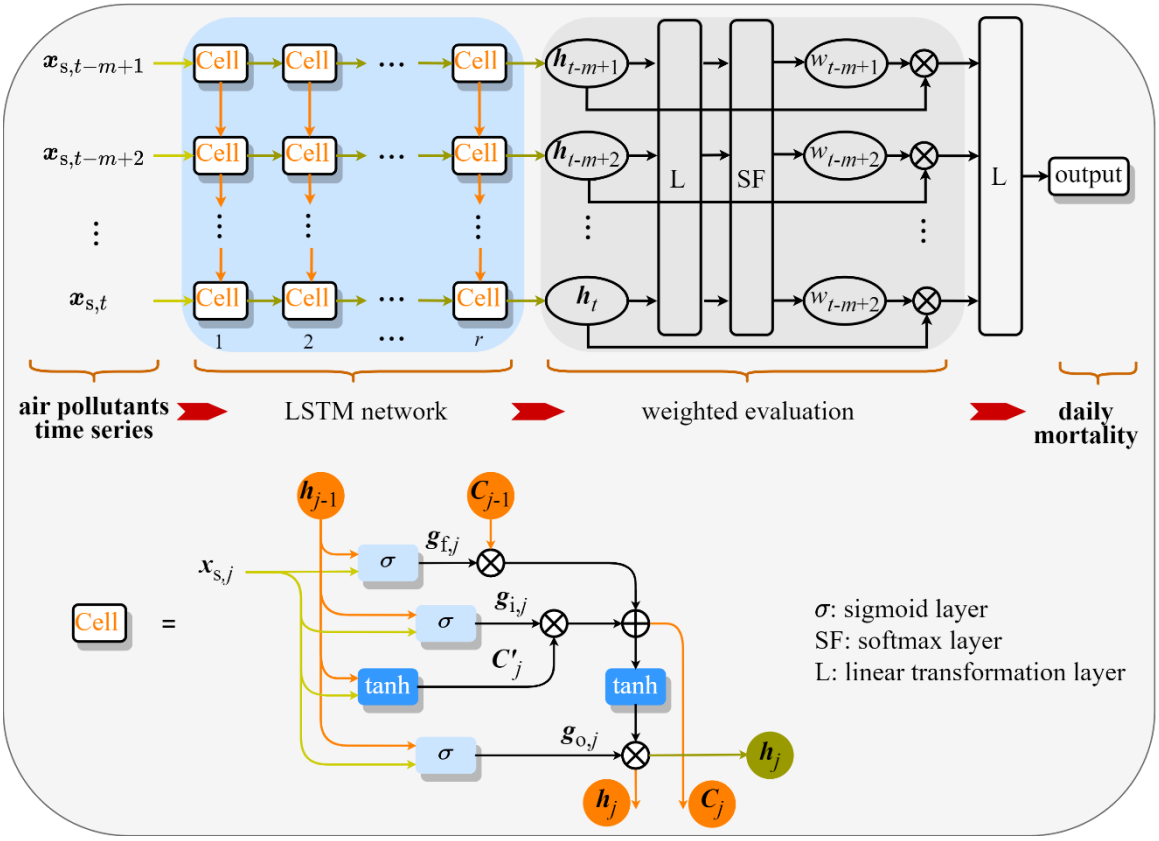

Fig. 1: Structure of LSTM based Air Pollutants-Related Health Consequence Assessment Model.

For evaluation period t, the input to the model is an air pollutants exposure series of previous m periods as [$\boldsymbol{x}_{s,t-m+1}$, $\boldsymbol{x}_{s,t-m+2}$,…, $\boldsymbol{x}_{s,t}$]. Note that each element of the sequence is a vector form for multiple air pollutants input. The LSTM network has r recurrent layers to process the m-period input and each layer has l output hidden features. The LSTM cell mainly consists of three specifically designed gates for information flow manipulation as shown in Fig. 1. In order to simplify the expression, layer number in following model description is neglected. Gate is essentially a bitwise multiplication operation after transformation using sigmoid function as σ. For period j, j=t-m+1, t-m+2, …, t, the coefficient $\boldsymbol{g}_{f,j}$ from the forget gate is expressed as Eqn. (3),

$$\boldsymbol{g}_{f,j} = \sigma(\boldsymbol{w}_f \boldsymbol{x}_{s,j} + \boldsymbol{u}_f \boldsymbol{h}_{j-1} + \boldsymbol{b}_f) \tag{3}$$

where $\boldsymbol{w}_f$, $\boldsymbol{u}_f$ and $\mathbf{b}_f$ are weights and bias parameters. The value of $\boldsymbol{g}_{f,j}$ is determined by the combination of exposure in j-th period as $\boldsymbol{x}_{s,j}$ and hidden layer output $\boldsymbol{h}_{j-1}$ from *(j-1)*-th period. This forget gate coefficient determines how much information from exposure in *(j-1)*-th period will be used in j-th period. The input gate coefficient $\boldsymbol{g}_{i,j}$ and new exposure information input $\boldsymbol{C'}_j$ from the j-th period are as Eqn. (4) and Eqn. (5) respectively,

$$\boldsymbol{g}_{\mathrm{i},j} = \sigma(\boldsymbol{w}_{\mathrm{i}}\boldsymbol{x}_{\mathrm{s},j} + \boldsymbol{u}_{\mathrm{i}}\boldsymbol{h}_{j-1} + \boldsymbol{b}_{\mathrm{i}}) \quad (4)$$

$$\boldsymbol{C'}_j = \tanh(\boldsymbol{w}_{\mathrm{c}}\boldsymbol{h}_{j-1} + \boldsymbol{u}_{\mathrm{c}}\boldsymbol{x}_{\mathrm{s},j} + \boldsymbol{b}_{\mathrm{c}}) \quad (5)$$

where $\boldsymbol{w}_{\mathrm{i}}$, $\boldsymbol{u}_{\mathrm{i}}$, $\boldsymbol{w}_{\mathrm{c}}$, and $\boldsymbol{u}_{\mathrm{c}}$ are weights and $\boldsymbol{b}_{\mathrm{i}}$ and $\boldsymbol{b}_{\mathrm{c}}$ are biases. With forget gate coefficient and input gate coefficient, the state of the j-th cell is updated using Eqn. (6),

$$\boldsymbol{C}_j = \boldsymbol{g}_{\mathrm{f},j} \odot \boldsymbol{C}_{j-1} + \boldsymbol{g}_{\mathrm{i},j} \odot \boldsymbol{C'}_j \quad (6)$$

where $\odot$ is the Hadamard product. The state $\boldsymbol{C}_j$ of j-th period filters out partial information from last period and adds new information from current period. After cell state update, the output gate parameter $\boldsymbol{g}_{\mathrm{o},j}$ and the cell output $\boldsymbol{h}_j$ are formulated as Eqn. (7) and Eqn. (8),

$$\boldsymbol{g}_{\mathrm{o},j} = \sigma(\boldsymbol{w}_{\mathrm{o}}\boldsymbol{x}_{\mathrm{s},j} + \boldsymbol{u}_{\mathrm{o}}\boldsymbol{h}_{j-1} + \boldsymbol{b}_{\mathrm{o}}) \quad (7)$$

$$\boldsymbol{h}_j = \boldsymbol{g}_{\mathrm{o},j}\tanh(\boldsymbol{C}_j) \quad (8)$$

where $\boldsymbol{w}_{\mathrm{o}}$, $\boldsymbol{u}_{\mathrm{o}}$ and $\mathbf{b}_{\mathrm{o}}$ are the output gate parameters. The output of a cell is determined by current cell state $\boldsymbol{C}_j$, output of last period $\boldsymbol{h}_{j\text{-}1}$, and current input $\boldsymbol{x}_{\mathrm{s},j}$. Except for the first layer, the inputs of other layers are the cell outputs of previous corresponding layers. For the LSTM framework with l layers in Fig. 1, its final outputs are $\boldsymbol{h}_{t\text{-}m+1}$, $\boldsymbol{h}_{t\text{-}m+2}$, …, $\boldsymbol{h}_t$ that contain dependent information from the input air pollutants exposure series. As these outputs from LSTM may have different contributions to the health consequences at time period t of interest, a weighted evaluation of their influence is further made using a softmax layer after a linear combination layer and the influence of each period k on the consequence is calculated as Eqn. (9),

$$w_k = \frac{\exp(L(\boldsymbol{h}_k))}{\sum_k \exp(L(\boldsymbol{h}_k))}, k = t-m+1, t-m+2, \ldots, t. \quad (9)$$

where $\boldsymbol{w}_k$ is the weight used to evaluate the contribution of k-th output from the LSTM to the health outcome at t and L is a linear combination layer. In this formulation, all the historical information of m periods before time period t is used to generate the health outcome at t. At last, the health outcome of period t is predicted through a weighted sum of LSTM outputs using a linear output as Eqn. (10),

$$output_t = L(w_k\boldsymbol{h}_k), k = t-m+1, t-m+2, \ldots, t. \quad (10)$$

where L means linear out and the output is public health outcome at period t (e.g., mortality count, morbidity count, cardiovascular disease count, etc.).

Note that the above formulation of a LSTM based health effect assessment model, the LSTM framework is not only used to process and extract chronically dependent information from the input sequenced air pollutants exposure from every single layer, but also used to capture and fit the relation between the multiple air pollutants exposure and the health consequence with its multiple layers as well as the subsequent weighted evaluation layer and linear output layer.

4. Experimental Results

To evaluate the performance of the above model in assessment of public health effects, air pollutants and related non-accidental mortality data from Chicago from Jan. 1, 1987 to Dec. 31, 2000 are used for experimentation. The selected inputs for the model are daily mean concentration level of PM_{10} and O_3, as well as daily mean temperature, which are frequently used in relevant studies [6, 10, 11, 21], being a classic and easily accessible dataset. Each category of data is pre-processed with interpolation, standardization, and reorganization. Then the pre-processed datasets are split into two parts, 70% for training the proposed model and the remainder (30%) for evaluating its performance. Performance metrics used are RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) as Eqns. (11) and (12), with RMSE evaluating the deviation between predicted mortality and true mortality and MAE evaluating absolute prediction error. Formally,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{P},i} - y_{\mathrm{T},i})^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{\mathrm{P},i} - y_{\mathrm{T},i}| \tag{12}$$

where $y_{\mathrm{P},i}$ and $y_{\mathrm{T},i}$ are the i-th predicted value and true value, respectively. n is the number of samples used. The experiments are performed using PyTorch version 2.1.1+cu121 [24]. The main parameters of the model used in the experiments are as follows. The numbers of hidden features (l) and recurrent layers (r) are 13 and 5, respectively. A variable learning rate is used with the initial value set to 0.05 and a decaying rate of 0.95 every 100 training epochs. The number of total training epochs is set to 8000.

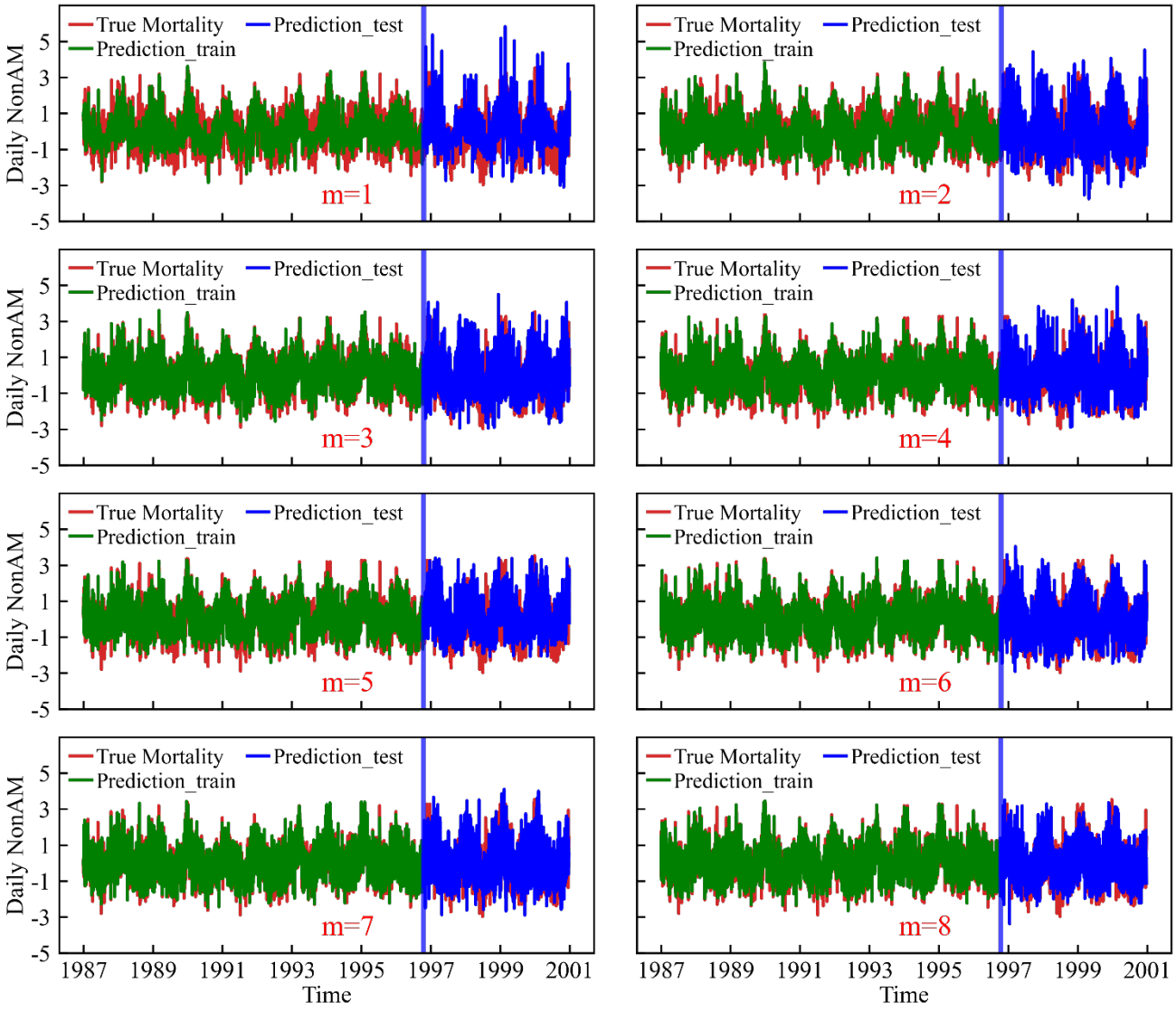

Fig. 2: Daily Mortality Prediction Comparison with Different m (Daily NonAM: daily non-accidental mortality, Prediction_test: prediction on testing set, Prediction_train: prediction on training set).

Table 1: Performance Metrics with Different *m* on Training and Testing Sets.

Data sets	Metrics	*m*=1	*m*=2	*m*=3	*m*=4	*m*=5	*m*=6	*m*=7	*m*=8
Training set	RMSE	0.6493	0.0666	0.0613	0.0016	0.0000	0.0007	0.0019	0.0104
	MAE	0.4485	0.0167	0.0171	0.0005	0.0000	0.0002	0.0006	0.0029
Testing set	RMSE	1.5313	1.8618	1.8588	1.4897	1.3949	1.4564	1.4050	1.5198
	MAE	1.1645	1.4640	1.4362	1.1828	1.1171	1.1667	1.1201	1.2028

With PM_{10}, O_3 and temperature as the predictors, the performance of the proposed health effect model with input series length from *m*=1 to *m*=8 are compared and experimental results are shown in Fig. 2 and Table 1, where all results are evaluated on the standardized datasets and *m* is the number of days of air pollutants exposure before the day when health consequence of interest (mortality) occurs. It is obvious that all eight experiments capture significant portions of the fluctuations of the mortality on both the training and testing sets, which shows that the proposed model has potential in assessing air pollutant-related public health effects, from a pure modelling perspective: the model has good forward predictive performance. For the training set, both metrics have downward trends with *m* increasing from 1 to 8. Part of the reason for this change is that a longer input series means more information are provided to the network, where more useful features and information are extracted for assessing the daily mortality. However, this downward trend can hardly be seen for the testing set. The first reason comes from the noise of training set, where a larger *m* brings both more useful information as well as higher data noise. The model learns mortality-related information and noise at the same time, which may lead to overfitting. The second one is that data distribution of the testing set may not be exactly the same as that of the training set, which lowers the generalization ability of the model. Although performance on the testing set are not perfect, the proposed model still shows great potential in the air pollutant-related public health effect problems. Explaining the health response sufficiently using the air pollutants as predictors is, after all, the structure of the classic models used in the field. Typically, this is done in-frame, predicting all points simultaneously, with almost no out-of-band prediction potential. This test shows that LSTMs have the potential to forecast such problems using nonlinear constructions from multiple air pollutants simultaneously.

There are several points worth noting for the proposed model in the experiments. First, model parameters keep the same for eight experiments, whereas parameters for different *m* can further be finely tuned for performance improvement. Second, the aforementioned downward trends for both metrics may not be absolute for longer *m* as a longer input air pollutant series may bring redundant information that exceeds the feature extraction and relation capture ability of the model. Third, de-noising can be performed on the original air pollutants and health consequence datasets to improve the performance of the model. In addition, diversified sampling methods can be used to split the training and testing sets to reduce the influence from variation of data distribution and feature selection for the inputs can help to find the most relevant influencing air pollutants. These measures will further improve the performance of the proposed model.

5. Conclusion

Ambient air pollutant-related public health effects have been a research hotspot for years in environmental epidemiology. For the extensively used assessment models like GAMs, there exist challenges in dealing with accumulated impacts of air pollutants exposure from distributed lags for fitting reasons, as well as the cross impacts from different pollutants, despite efforts such as [22, 23]. In this paper, an LSTM network-based model is proposed to evaluate the impacts from distributed lags of air pollutants exposure on public health outcomes. The LSTM framework is used to extract the outcome-related features from the exposure series and then a weighted evaluation is used to assess their impacts the health outcome. Air pollutants and non-accidental mortality data of Chicago from NMMAPS program are used to test the performance of the proposed method and experimental results show that the proposed model has potential in capturing the accumulated impacts from the exposure with distributed lags and dealing with cross impacts from multiple air pollutants. In the future work, de-noising of the datasets, feature selection of multiple air pollutants and finer tuning of model hyper-parameters can be performed to further improve the performance of the proposed model, especially for evaluating a specific health consequence.

The primary drawback to this approach is interpretability, as it is difficult to summarize the large numbers of coefficients that result from such neural network models. This is an open area of research and will require significant further effort to allow for comparability of this new approach with previously published and validated frameworks.

References

[1] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," *Nature*, vol. 52.5, no. 7569, p. 367–371, 2015.

[2] C. A. Pope Iii, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *Journal of the American Statistical Association*, vol. 287, no. 9, pp. 1132–1141, 2002.

[3] F. Dominici, A. McDermott, S. L. Zeger, and J. M. Samet, "On the use of generalized additive models in time-series studies of air pollution and health," *American Journal of Epidemiology*, vol. 156, no.3, pp.193–203, 2002.

[4] F. Dominici, A. McDermott, T. J. Hastie, "Improved semiparametric time series models of air pollution and mortality," *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 938–948, 2004.

[5] J. B. Souza, A. R. Valdério, C. F. Glaura, I. Márton, B. Pascal, and J. M. Santos, "Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data," *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, vol. 67, no. 2, pp. 453-480, 2018.

[6] W. S. Burr, G. Takahara, H. H. Shin, "Bias correction in estimation of public health risk attributable to short-term air pollution exposure," *Environmetrics*, vol. 26, no. 4, pp. 298-311, 2015.

[7] W. S. Burr, "Air pollution and health: Time series tools and analysis," Ph.D. dissertation, Dept. of Mathematics & Statistics, Queen's Univ., Kingston, ON, Canada, 2012.

[8] F. Dominici, R. D. Peng, C. D. Barr, and M. L. Bell, "Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach," *Epidemiology*, vol. 21, no. 2, pp. 187-194, 2010.

[9] T. Wei, "Associations between short-term exposure to ambient air pollution and lung function in adults," *Journal of Exposure Science & Environmental Epidemiology*, pp. 1-9, 2023.

[10] S. Jarvis and W. S. Burr, "Development of a multi pollutant model to assess air pollution association with human health effects," in *Proceedings of the* 4th *International Conference on Statistics*: *Theory and Applications* (*ICSTA*'22), Prague, Czech Republic, 2022, DOI: 10.11159/icsta22.151.

[11] H. H. Shin, J. Owen, A. Maquiling, R. P. Parajuli and M. Smith-Doiron, "Circulatory health risks from additive multi-pollutant models: short-term exposure to three common air pollutants in Canada," *Environmental Science and Pollution Research*, vol. 30, no. 6, pp. 15740-15755, 2023.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[13] K. M. Tarwani and S. Edem, "Survey on recurrent neural network in natural language processing," *International Journal of Engineering Trends Technology*, vol. 48, no. 6, pp. 301-304, 2017.

[14] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," *IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 273-278, 2013.

[15] B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, "Forecasting air quality time series using deep learning," *Journal of the Air & Waste Management Association*, vol. 68, no. 8, pp. 866-886, 2018.

[16] R. D. Peng and L. J. Welty, "The NMMAPSdata package," *R News*, vol. 4, no. 2, pp. 10–14, 2004.

[17] S. Castel and W. S. Burr, "Assessing statistical performance of time series interpolators," *Engineering Proceedings*, vol. 5, no. 1, pp. 1-11, 2021, DOI: doi.org/10.3390/engproc2021005057.

[18] A. Pandey and A. Jain, "Comparative analysis of KNN algorithm using various normalization techniques," *International Journal of Computer Network and Information Security*, vol. 9, pp. 36-42, 2017.

[19] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp.2554-2558, 1982.

[20] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, 2016.

[21] R. D. Peng and F. Dominici, Statistical methods for environmental epidemiology with R: a case study in air pollution and health. Springer, 2008.

[22] A. Gasparrini, B. Armstrong, and M.G. Kenward, 2010. "Distributed lag non-linear models". *Statistics in Medicine*, 29(21), pp.2224-2234.

[23] W.S. Burr, H.H. Shin, and G. Takahara, 2019. “Synthetically lagged models”. *Statistics & Probability Letters*, 144, pp.37-43.

[24] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga and A. Desmaison, "Pytorch: An imperative style, high-performance deep learning library." *Advances in neural information processing systems* 32 (2019).